IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Appln No.:	09/939,392
Applicants:	James Roberge et al.
Filed:	August 24, 2001
For:	Structured Speech Recognition
TC/A.U.:	2655
Examiner:	Huyen X. Vo
Docket No.:	71416
Customer No.:	22242

Confirmation No. 3939

CERTIFICATE OF MAILING

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on this date.

9/20/05
Date

Steven G. Parmelee
Registration No. 28,790
Attorney for Applicant(s)

APPEAL BRIEF

Mail Stop APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Pursuant to 37 C.F.R. §1.192, the applicants hereby respectfully submit the following Brief in support of their appeal. Pursuant to 37 C.F.R. §1.192(a) this brief is being filed in triplicate.

(1) Real Party in Interest

The real party in interest is Cyberpulse, a limited liability corporation having a primary place of business in Highland Park, Illinois.

(2) Related Appeals and Interferences

There are no related appeals or interferences known to appellant, the appellant's legal representative, or assignee that will directly affect, or be directly affected by or have a bearing on the Board's decision in the pending appeal.

(3) Status of Claims

Claims 1 - 15 are pending and presently stand twice and finally rejected and constitute the subject matter of this appeal.

(4) Status of Amendments

No post-final amendments have been submitted.

(5) Summary of Claimed Subject Matter

The applicant's present a structure speech recognition apparatus and method having a preferred (though not exclusive) application as a data-entry facilitator for a medical records database using, in particular, a hierarchically-organized database representation. Fig. 1, reproduced below, depicts a speech preprocessor that organizes information to be used at run-time by, ultimately, a speech engine that receives utterances by a user.[1]

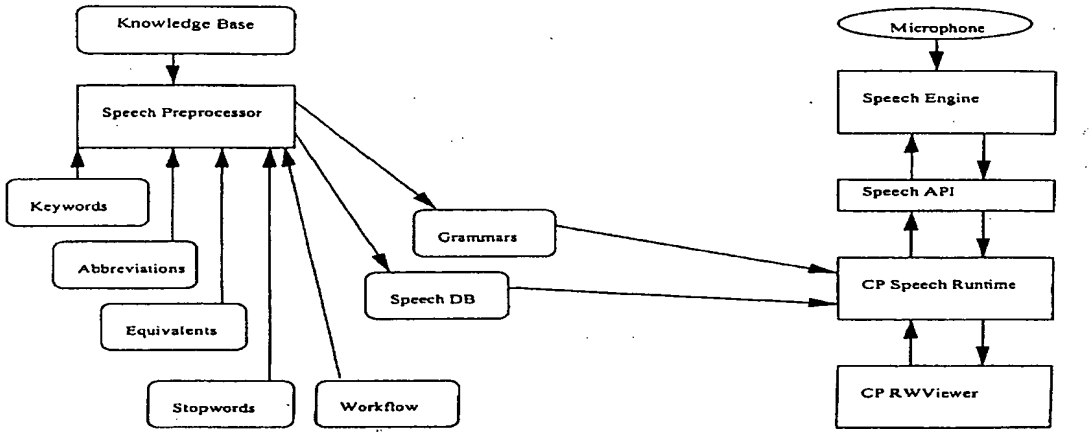


Figure 1

At least part of this information to be organized comprises a workflow function (exemplified herein as contexts for populating selected data items in a main database). The workflow function specifies groups of findings (such as medical findings) that are grouped to enhance workflow, to establish a vocal context, and to establish a set of navigational

1 Application page 4, lines 4-5.

commands. There can be many such workflows to accommodate the needs and styles of different target users. For example, the workflow for a cardiologist can be helpfully based on anatomical structure (for example, left ventricle, mitral valve, and so forth) while the workflow for a radiologist can be based instead upon pathology (for example, lesion type, and so forth).[2]

As noted above, these teachings are suitable for use with a knowledge base that is represented as a hierarchically-organized database. An illustrative representation appears in FIG. 2 (reproduced below) that provides the echocardiogram aspects of a medical knowledge base where breaks occur as between normal cardiac structures, comparisons to prior studies, and conclusions (with conclusions being further segmented into data fields of the database as associated with the hierarchical knowledge base).[3]

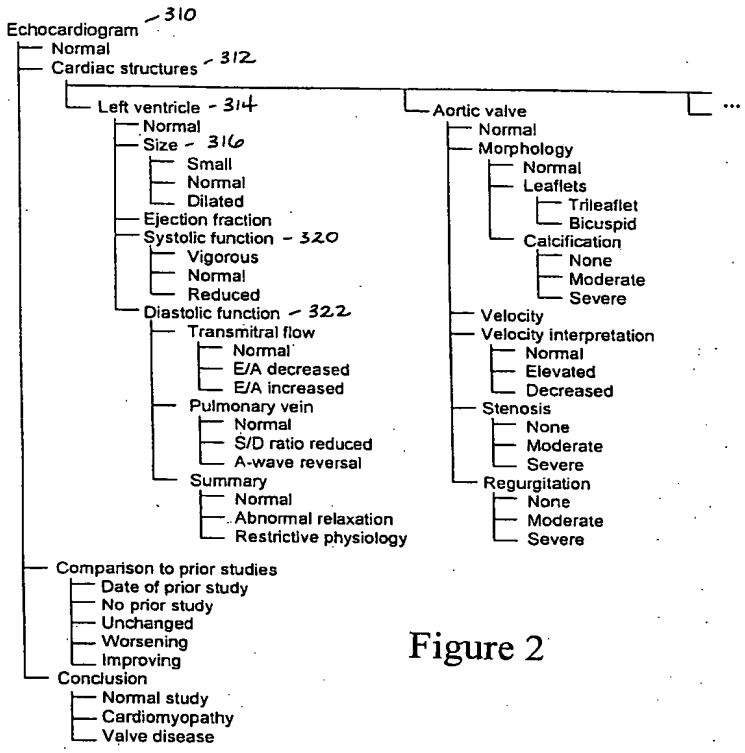


Figure 2

2 Application page 4, lines 8-17.
3 Application page 5, line 17-page 6, line 31.

To illustrate, the applicant teaches that "Navigation for data entry in the user interface is done by verbally selecting a structure of interest or finding a structure of interest in the menu system or summary viewer . . . ."[4] As noted above, a set of navigational commands are established for each workflow function (and, accordingly, for a corresponding vocal context). These navigational commands permit a user to navigate, using spoken commands, about a displayed user interface such as that depicted in FIG. 4 (reproduced below).

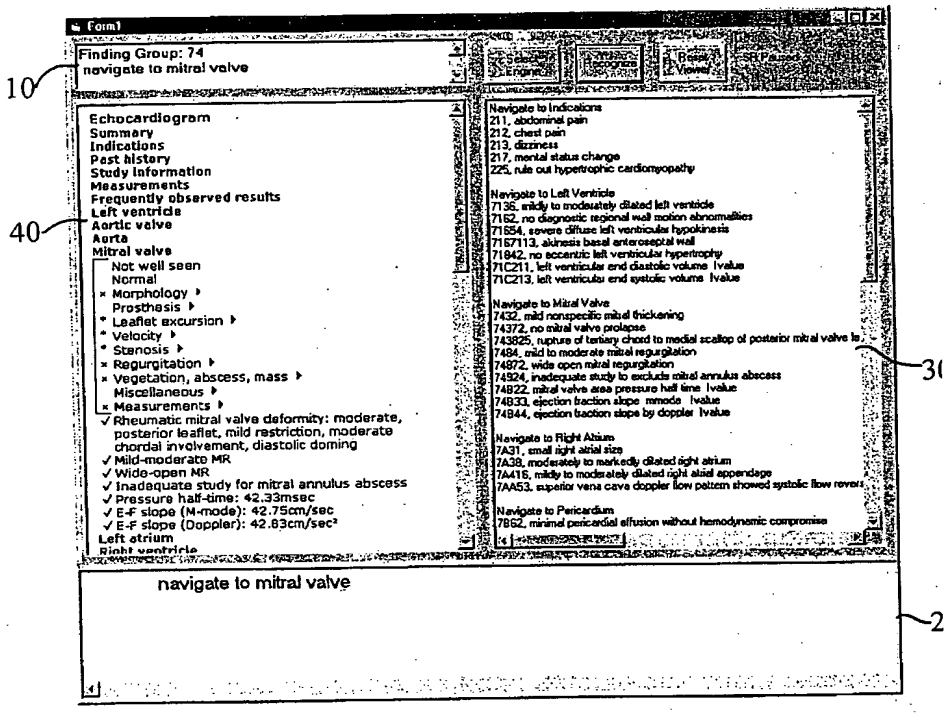


Figure 4

In this illustrative user interface, the screen snapshot depicts a small panel (10) at the upper left that contains current workflow context and a larger panel (20) at the bottom that contains returned commands to process or find data items to record in the main database.

The present application presumes that the person providing the speech to be recognized is present and actively interacting with the input mechanism itself. This difference becomes clearly manifest in the applicant's requirement that this person be able to interact with the process by providing spoken commands to that process. In particular, the

4 Page 6, lines 22-24.

applicant provides the user with an ability to navigate the data entry process by accommodating navigational commands.

To illustrate, the applicant teaches that "Navigation for data entry in the user interface is done by verbally selecting a structure of interest or finding a structure of interest in the menu system or summary viewer . . . ."In this illustration, the user has uttered the navigational command, "Navigate to mitral valve." The system has recognized this navigational command as an appropriate command and has executed that command to set the current finding group to "mitral valve" (as reflected by the contents of the current workflow context panel (10). In a preferred embodiment and as illustrated here, each such finding group has its own word-mapping database and corresponding grammar for the speech recognition engine to help improve recognition accuracy.[5]

To complete the example, and with reference to FIG. 5 (reproduced below), when the user then subsequently utters the phrase, "moderate to severe stenosis," a check mark is placed by "moderate-severe" under the "mitral valve – stenosis" outline, to indicate successful entry of this finding. In addition, the corresponding narrative finding "moderate to severe mitral stenosis" appears in the bottom panel as well (note that this narrative expression occurs notwithstanding the lack of the word "mitral" in the originally uttered phrase "moderate to severe stenosis" due, at least in part, to the user having first navigated to the mitral value context via the earlier entered navigational command).[6]

So configured, a user can, if he wishes, utter a navigational command to identify, for example, a particular context of interest and then utter a finding to be entered. That latter utterance is recognized and disambiguated as a function, in part, of the context established in part by the navigational command.

(6) Grounds of Rejection to be Reviewed on Appeal

Claims 1, 3, 4, 8, 9, and 11-15 are rejected under 35 U.S.C. 102(b) given Ryan (U.S. Patent No. 5,809,476) ("Ryan"). Claims 2, 6, and 10 are rejected under 35 U.S.C. 103(a) given Ryan in view of Zhilyaev (U.S. Patent No. 6,137,911) ("Zhilyaev"). Claims 5 and 7

5 Application page 9, line 30-page 10, line 5.
6 Application page 10, lines 8-16.

are rejected under 35 U.S.C. 103(a) given Ryan in view of Zhilyaev and further in view of Li (U.S. Patent No. 5,774,588) ("Li"). The applicant disputes these rejections.

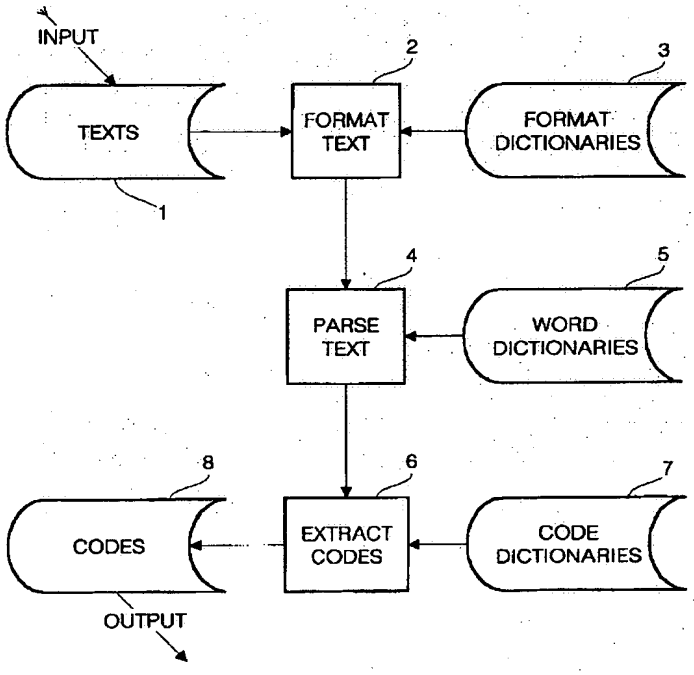
(7) **Argument**

Rejections under 35 U.S.C. 102(b)

Claims 1, 3, 4, 8, 9, and 11-15 are rejected under 35 U.S.C. 102(b) given Ryan (U.S. Patent No. 5,809,476) ("Ryan").

As all of the claims have been rejected with Ryan serving as the only or as a primary reference the applicant believes it would be helpful to first briefly describe and characterize Ryan's teachings.

Ryan provides a system to convert information, such as medical information, into corresponding representative codes. For example, the words "femur at the knee" can be represented, according to Ryan, by the corresponding code "!B@C." Ryan's FIG. 1 provides a general overview of his teachings (with FIG. 1 being reproduced below for the convenience of the reader).



Ryan begins with inputting narrative text (gleaned, he suggests, from keyboard entry, speech recognition, and so forth). Ryan suggests no constraints with respect to the nature or form of that narrative entry nor does Ryan offer any thoughts with respect to conducting speech recognition in a successful manner. Such details are left to the imagination of the reader. Instead, Ryan presumes provision of accurate narrative text, however obtained.

Ryan then formats that narrative text to remove punctuation marks and plural or tense modifiers. Format dictionaries (3) are used to facilitate this formatting. The resultant formatted narrative text is then parsed (at block 4) to separate each individual word to permit comparison of each parsed word with available dictionary entries (5). Some pertinent additional details regarding this process appear in FIG. 2 (also reproduced below for the convenience of the reader).

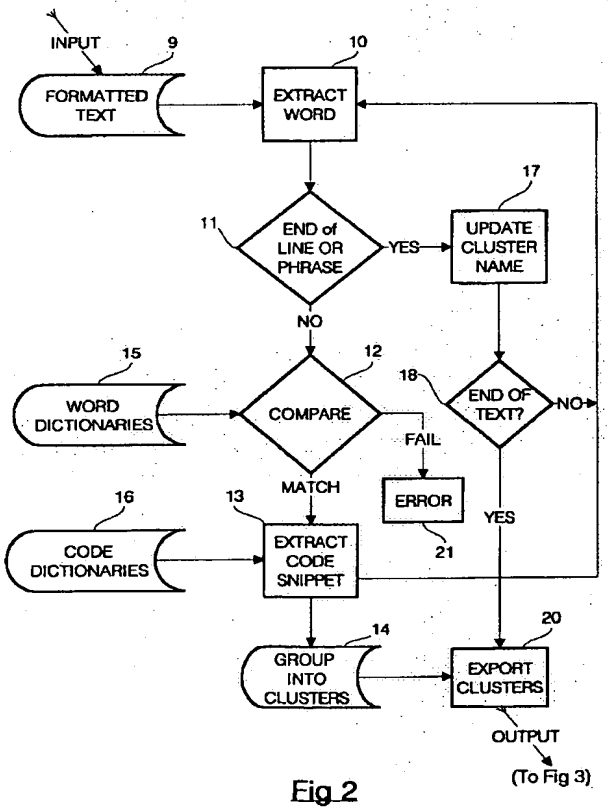


Fig 2

In particular, Ryan compares (12) each parsed word against the entries in his word dictionaries (15). When no match occurs, an error condition (21) results. When a match does occur, the corresponding code for that word is "extracted" from a code dictionary (16).[7]

Ryan then proceeds to group his extracted codes into related groups (to capture whole thoughts or expressions). Ryan also then examines each such group, referred to by Ryan as a "cluster," to ensure that the contents of that group are complete. In particular, Ryan requires that each such group comprise at least one code from a group of codes as are used for anatomical modifiers, at least one code from a group of codes as are used for clinical modifiers, and at least one code from a group of codes as are used for "side information."[8] When a group is found to be incomplete, Ryan provides some suggestions for how the content requirements might be met.

Ryan is therefore seen to teach a translation mechanism that begins with narrative text and that outputs a completely coded translated output, where that translated output is comprised of codes that are themselves comprised of the codes that represent the individual parsed words of the original text. It is Ryan's intent and purpose to convert and translate an original narrative sentence such as, "This patient requires a left to right shunt" into a code such as, "AxL/R$."

At the outset, it may be observed that Ryan's concerns and approach are considerably different from those addressed and set forth in the present application. Ryan is unconcerned with *how* he obtains his narrative text, for example, whereas in the present case concerns regarding the accuracy, speed, and ease-of-use that typically surround speech recognition are important issues. This difference, alone, leads to numerous operational differences between the present claims and Ryan's teachings.

For the sake of brevity, however, the applicant is content here to note one particular significant difference between the present application and Ryan. This difference stems from the fact that, while Ryan is unconcerned with the source of his original narrative text[9] the present application presumes that the person providing the speech to be recognized is present

7 In fact, at this point, Ryan tends to refer to the code as a "code snippet," as he anticipates combining this code snippet with other related code snippets to form, in their aggregate, a resultant code for a plurality of related words as comprise, for example, a phrase, statement, or sentence.
8 Where "side information" refers to a side of the human body as in, for example, "right" or "left" side.

and actively interacting with the input mechanism itself. This difference becomes clearly manifest in the applicant's requirement that this person be able to interact with the process by providing spoken commands to that process. In particular, the applicant provides the user with an ability to navigate the data entry process by accommodating navigational commands.

To illustrate, the applicant teaches that "Navigation for data entry in the user interface is done by verbally selecting a structure of interest or finding a structure of interest in the menu system or summary viewer . . . ."[10] As a further illustration, the applicant provides an example wherein, "At this point the phrase 'Navigate to mitral vale' is uttered by the user. The results are illustrated in FIG. 4. The system recognized that 'navigate to mitral valve' was the appropriate command (indicated in the large panel 20 at the bottom), and executed that command to set the current Finding Group to Mitral Valve."[11] To support this active role on the part of the user, the applicant provides for navigational commands that correspond to particular contexts and workflows. Ryan makes no such provision.

For example, claim 1 provides for "developing a series of contexts, each context comprising a series of *navigational commands* for populating a selected series of data items of the main database . . . ."[12] Ryan makes no provision for commands of any type and in particular has no navigational commands as described.

Somewhat similarly, independent claim 8 provides for, "defining a workflow for populating a selected set of data items in the main database, each workflow providing custom knowledge comprising a series of *navigational commands* and a context identifier . . . ."[13] and independent claim 13 provides for, "a context identifier for establishing a workflow for information processing of a series of *navigational commands* for populating a selected series of data items in the main database . . . ." Again, Ryan makes no provision for commands of this sort, let along navigational commands that correlate as described to a specific workflow context.

---

9 To the point, in fact, that the originator of the narrative text need not even be present or aware of Ryan's code translation processing.
10 Application page 6, lines 22-24.
11 Application page 9, lines 30-34.
12 Emphasis provided.
13 Emphasis provided.

The applicant therefore respectfully submits that Ryan cannot be said to anticipate the recitations of these claims. Therefore, independent claims 1, 8, and 13 are distinguished from Ryan and may be passed to allowance.

The remaining claims are ultimately dependent upon one of these independent claims. In addition, these claims introduce additional content that, particularly when viewed in context with the claim or claims from which they depend, constitutes additional incremental patentable subject matter. For all these reasons the applicant respectfully submits that these dependent claims are in suitable form to support allowance.

Rejections under 35 U.S.C. 103(a)

Claims 2, 6, and 10 are rejected under 35 U.S.C. 103(a) given Ryan in view of Zhilyaev (U.S. Patent No. 6,137,911) ("Zhilyaev"). Claims 5 and 7 are rejected under 35 U.S.C. 103(a) given Ryan in view of Zhilyaev and further in view of Li (U.S. Patent No. 5,774,588) ("Li").

These claims are ultimately dependent upon one of these independent claims. In addition, these claims introduce additional content that, particularly when viewed in context with the claim or claims from which they depend, constitutes additional incremental patentable subject matter. For all these reasons the applicant respectfully submits that these dependent claims are in suitable form to support allowance.

(8) Claims Appendix

1. (Original) A method for populating a main database from speech recognition output based on verbal utterances of a user, said method comprising:

a) developing a series of contexts, each context comprising a series of navigational commands for populating a selected series of data items of the main database, and each context represented by a context identifier;

b) creating a word-mapping database for each context, said word-mapping database containing a words drawn from narrative statements (written and oral) associated with the data items in the selected context, as well as linkages between these words and data items;

c) identifying a selected context by comparing the context identifier of the selected context with speech recognition output generated based on a context-identifying verbal utterance of the user;

d) recording selected data items within the selected context by mapping the speech recognition output generated based on utterances of the user to data items in the main database using the word-mapping database for the selected context; and

e) repeating steps c) and d) until the user finishes entering data, thereby populating the main database.

2. (Original) The method of claim 1, wherein the series of contexts and word-mapping databases are developed using a hierarchically-organized database representation based on knowledge regarding the relationship of data items in the main database, said hierarchically-organized database representation having a plurality of nodes capable of having further related nodes, fields, or attributes.

3. (Original) The method of claim 1, wherein the main database is a medical records database and the series of contexts are developed based on completion of data entry for generation of a medical report.

4. (Original) The method of claim 1, wherein speech recognition output is mapped to data items in the selected context using a word-mapping database that includes a set of designated keywords representing spoken phrases for populating data items and by comparing speech recognition output to these keywords.

5. (Original) The method of claim 4, wherein speech recognition output is compared to the words of the word-mapping database by:

a) calculating the number of keywords in a phrase that match speech recognition output to generate a keyword match number;

b) prioritizing phrases with identical keyword match numbers by calculating the ratio of keywords in a phrase that match speech recognition output to the total keywords in a phrase to generate a keyword match ratio; and

c) prioritizing phrases with identical keyword match ratios by performing a string comparison algorithm.

6. (Original) The method of claim 4, wherein the word mapping is performed by calculating the number of keywords in a phrase that match speech recognition output, by calculating the ratio of keywords in a phrase that match speech recognition output to the total keywords in a phrase, or by performing a string comparison algorithm.

7. (Original) The method of claim 6, wherein the string comparison algorithm comprises the Edit Distance method.

8. (Original) A method for populating a main database from speech recognition output based on verbal utterances of a user, comprising:

a) defining a workflow for populating a selected set of data items in the main database, each workflow providing custom knowledge comprising a series of navigational commands and a context identifier for associating spoken utterances with explicitly coded data relating to the workflows;

b) identifying the workflow being employed;

c) creating a word-mapping database for each workflow; and

d) populating the selected series of data items for the identified workflow using the series of navigational commands and comparing speech recognition output generated based on the verbal utterances of the user to obtain words and explicitly coded data for populating the main database.

9. (Original) A method as recited in claim 8, wherein the defining of the workflows comprises developing a series of contexts for populating the selected data items in the main database with each context being represented by the context identifier.

10. (Original) A method as recited in claim 9, wherein the series of contexts are developed using a hierarchically-organized database representation.

11. (Original) A method as recited in claim 8, wherein the defined workflow comprises a medical records database and a series of context identifiers developed for the completion of a medical report.

12. (Original) A method as recited in claim 8, wherein the defined workflows use the context identifier to provide metrics for scoring the selected context's speech recognition output based on the context-identifying verbal utterance of the user.

13. (Original) A system for populating a main database from speech recognition output based on verbal utterances of the user, comprising:

a context identifier for establishing a workflow for information processing of a series of navigational commands for populating a selected series of data items in the main database;

a word-mapping database created for each context identifier comprising words of possible entries of the data items in the context;

a processor for comparing information from the context identifier with speech recognition output generated based on a context-identifying utterance of the user;

a memory device associated with the main database for populating the selected series of data items for the selected context of the context identifier using the series of navigational commands and the speech recognition output.

14. (Original) A system as recited in claim 13, wherein processor identifies the selected context associated with the context identifier using metrics for comparing the context identifier with the speech recognition output for generating scores associated with the context-identifying verbal utterance of the user.

15. (Original) A system as recited in claim 14, wherein said processor employs multi-level scoring for generating the speech recognition output.

(9) Evidence Appendix

Not applicable.

(10) Related Proceeding Appendix

Not applicable.

Respectfully submitted,

By: _____
Steven G. Parmelee
Registration No. 28,790

Date: September 20, 2005

FITCH, EVEN, TABIN & FLANNERY
120 South LaSalle, Suite 1600
Chicago, Illinois 60603-3406
Telephone: (312) 577-7000
Facsimile: (312) 577-7007